D1-82-0459

BOEING SCIENTIFIC RESEARCH LABORATORIES

Minimizing Functionals on Normed-Linear Spaces

A. A. Goldstein

Mathematics Research

August 1965

ERRATA AND ADDITIONS

"Minimizing Functionals on Normed-Linear Spaces"

by

A. A. Goldstein

Page 3. In (b) of the theorem replace $L_p[0,1]$ by a uniformly convex Banach space.

Page 4. Line 7 from bottom. Assume E is uniformly convex. By [6] p. 113.

Page 8. Line 6. for $p > 1$.

Page 10. Insert after sentence of line 5: and f is uniformly F-differentiable on S if $p > 2$. The inequality $|a|^{2r} + |b|^{2r} -2|a|^r|b|^r[a,b]/|a||b| \leq |a|^{2r-2}(r^2 + 1)|a - b|^2$ where $|a| > |b|$, and a direct computation show that the F-derivative f' exists and is lipschitz continuous for all $p > 1$.

Page 11. Line 12. Delete paragraph beginning with "By the theorem of I..." Replace this with the following: Moreover, the space $\mathcal{L}_r$ is uniformly convex. This follows by a theorem of Smulian [12], which states that if the norm in a B-space is uniformly F-differentiable on the unit sphere, then the conjugate space is uniformly convex.

Page 13. Replace ref [6] by M. Day, Normed Linear Spaces, *Academic Press*, N. Y., 1962.

Add:

[12] V. Smulian. Sur la derivabilité de la normed dans l'espace de Banach, C. R. (Doklady), *Acad. Sci. U.R.S.S.*, Vol. 27, (1940), p. 643-648.

D1-82-0459

MINIMIZING FUNCTIONALS ON NORMED-LINEAR SPACES

by

A. A. Goldstein
University of Washington

Mathematical Note No. 419

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

August 1965

ABSTRACT

This paper extends results of [1], [2], of Goldstein, and [3] of Vainberg concerning steepest descent and related topics. An example is given taken from a simple rendezvous problem in control theory. The problem is one of minimizing a norm on an affine subspace. The problem here is solved in the primal. A solution in the dual is given by Neustadt [4].

I. GENERATION OF MINIMIZING SEQUENCES

Let E be a normed linear space (n. l. space), x_0 an arbitrary point of E and f a functional defined on E. Let S denote the level set $\{x \in E : f(x) \leq f(x_0)\}$ defined at an arbitrary fixed $x_0 \in E$. We denote by $f'(x)$ the Frechet or F-derivative of f at x. We call f *uniformly F-differentiable* on S if f is F-differentiable on S and if $\delta(\varepsilon)$ in the definition of the F-derivative is constant on S. The F-derivative of f at x will be denoted by $f'(x)$. If $g \in E^*$ the value of g at x will be denoted by $[g,x]$, and if $h \in E^{**}$ the value of h at $g \in E^*$, by $[h,g]$. Recall that if E and F are n. l. spaces and A is bounded linear operator from E to F, in short $A \in B(E,F)$, and if A is onto then A^{-1} exists and belongs to $B(F,E)$ if and only if for some $m > 0$ and all $x \in E$, $||Ax|| \geq m||x||$; and that $m||x|| \leq ||Ax|| \leq M||x||$ for all x in E implies that $M^{-1}||y|| \leq ||A^{-1}y|| \leq m^{-1}||y||$ for all y in F.

We observe that if E is a reflexive Banach Space, $A \in B(E,E^*)$ and $[Ax,x] \geq m||x||^2$ for all $x \in E$, then A is onto and thus has an inverse. For, on the contrary supposition, take $f^0 \notin M = \text{range } A$. Choose g in E^{**} such that $g(f^0) = \text{dist}(f^0,M) > 0$, $||g|| = 1$ and $g(f) = 0$ for all f in M. Take $\bar{g}$ in E so that $[g,f] = [f,\bar{g}]$ for all f in E^*. Then $0 = [g,Ax] = [Ax,\bar{g}]$ for all x in E.

Thus $[A\bar{g},\bar{g}] = 0$ while $||g|| = ||\bar{g}|| = 1$. Q. E. D.

Let ϕ denote a bounded map from S to E satisfying the two conditions $[f'(x),\phi(x)] \geq 0$, and given $\varepsilon > 0$ there exists $\delta > 0$ such that $[f'(x),\phi(x)] < \delta$ implies $||f'(x)|| < \varepsilon$. Some examples of such mappings are the following:

(1) Let $A \in B(E^*,E)$ such that $[y,Ay] \geq \sigma||y||^2$ for all $y \in E^*$ and some $\sigma > 0$. Let $\phi(x) = Af'(x)$ and choose $\delta = \varepsilon^2\sigma$. Then $||f'(x)|| < \varepsilon$.

As a possible candidate for the operator A, suppose f is twice F-differentiable on E. Assume that for some $\mu > 0$, and some x in S the operator $f''(x)$ in $B(E,E^*)$ is onto and is *"bounded below"*, that is, the bilinear functional satisfies $[f''(x)z,z] \geq \mu||z||^2$ for all z in E. Then $||f''(x)z|| \geq \mu||z||$ showing that $f''(x)$ has an inverse $[f''(x)]^{-1} = A \in B(E^*,E)$. Since A has a bounded inverse, there exists a number $\sigma > 0$ such that $||Ay|| \geq \sigma||y||$ for all $y \in E^*$. Set $z = Ay$. Then $[f''(x)z,z] = [y,Ay] \geq \mu\sigma^2||y||^2$ showing the candidacy of A.

(2) Suppose E is a reflexive Banach space. By the weak compactness of the unit sphere in E it follows that for some z_0, $||z_0|| = 1$, $[f'(x),z_0] = ||f'(x)||$. Set $\phi(x) = z_0||f'(x)||$. Because $[f'(x),\phi(x)] = ||f'(x)||^2$, $\phi(x)$ is the analogue of the gradient in Hilbert space. When E is an L_p space the point z_0 is obtained by considerations of equality in Hölder's inequality.

(3) Since $||f'(x)|| = \sup\{[f'(x),z]:||z|| = 1\}$, if $0 < \alpha < 1$ a point z_0 exists such $[f'(x),z_0] \geq \alpha||f'(x)||$. If for fixed α and all $x \in S$ we can find such z_0, we may take $\phi(x) = z_0||f'(x)||$.

In what follows let $\Delta(x,\rho) = f(x) - f(x - \rho\phi(x))$ and $g(x,\rho) = \Delta(x,\rho)/\rho[f'(x),\phi(x)]$. Assume E is a normed-linear space and S is the level set of f at x_0 in E. In what follows, assume $0 < \sigma < \frac{1}{2}$.

Theorem. Assume that on S f is uniformly F-differentiable or that the F-derivative f' exists and is uniformly continuous. Set $x_{k+1} = x_k$, when $[f'(x_k),\phi(x_k)] = 0$; otherwise choose* ρ_k so that $\sigma < g(x_k,\rho_k) \leq 1-\sigma$ when $g(x_k,1) < \sigma$ or $\rho_k = 1$ when $g(x_k,1) \geq \sigma$, and set $x_{k+1} = x_k - \rho_k\phi(x_k)$.

(a) If S is bounded or f is bounded below then $\{f'(x_k)\}$ converges to 0 while $\{f(x_k)\}$ converges downward to a limit, L. If S is compact, then every cluster point of $\{x_k\}$ is a zero of f'. In addition, if $\phi(x_k) \to 0$ and f' has finitely many zeros, $\{x_k\}$ converges.

(b) If S is convex and bounded and f is convex, $L = \inf\{f(x):x \in S\} = \theta$. If, in addition, E is a reflexive Banach space, then every weak cluster point of $\{x_k\}$ minimizes f on E. If $E = L_p[0,1]$, then $\{x_k\}$ converges to a unique minimizer of f.

(c) Assume that the Gateaux derivative f'' exists on S and satisfies $\mu||z||^2 \leq [f''(x)z,z] \leq M||z||^2$ for all $x \in S$, $z \in E$ and some $\mu > 0$. Assume S is convex and E is complete. Then $\{x_k\}$ converges to a unique minimizer of f on E.

---

*If the Gateaux differential f'' satisfies $f''(x,h,h) \leq ||h||^2/\rho_0$ for all h in E, x in S and some $\rho_0 > 0$ we can choose ρ_k to satisfy $\delta \leq \rho_k \leq 2\rho_0 - \delta$ with $0 < \delta \leq \rho_0$. The method of steepest descent could also be employed. See [9].

The proof of (a) is given in [1]. The proof there is stated for E, a Hilbert space, but the same proof works when E is taken to be a n. l. space. Two comments might be made, however. S bounded and f' uniformly continuous on S implies that f' is bounded on S. (See e.g. [5], p. 19.) It follows by employing the mean value theorem that f is bounded below on S. The statements f uniformly F-differentiable and the F-derivative f' is uniformly continuous are equivalent. (See [5], p. 45.)

(b) Given $\varepsilon > 0$ choose $z' \varepsilon E$ such that $f(z') = \theta + \varepsilon/2$. Because f' exists at x_k and f is convex, $f(z') \geq f(x_k) + [f'(x_k), z' - x_k]$. Since $\{f'(x_k)\} \to 0$ and S is bounded, for all k sufficiently large $f(x_k) \leq f(z') + \varepsilon/2 = \theta + \varepsilon$, showing that $L = \theta$.

If E is reflexive and S is convex, closed, and bounded, then S is weakly compact. Since f is convex, the sets $\{x \varepsilon E : f(x) \leq k\}$ are closed, convex, and weakly closed, for all k. Thus f is weakly lower semi-continuous. If z is a weak cluster point of $\{x_k\}$ then for an appropriate subsequence, $\lim\inf f(x_k) = L \geq f(z)$. Assume $E = L_p[0,1]$ and f is the norm on E. By [6], p. 78, if $\{x_k\}$ converges weakly to z and $f(x_k) \to z$ then $\{x_k\}$ converges strongly to z. It follows that every weak cluster point of $\{x_k\}$ is a strong cluster point of $\{x_k\}$. Since f' vanishes at every weak cluster point of $\{x_k\}$ and f' vanishes only once by the strict convexity of f, every subsequence of $\{x_k\}$ has the same weak cluster point z, showing that $\{x_k\}$ converges to z.

(c) The hypothesis of (c) imply that f' is Lipschitz continuous and that the set S is bounded. Otherwise S would contain an unbounded sequence, say $\{z_k\}$ By Taylor's theorem if $u \in S$, $f(z_k) \geq f(u) + ||z_k - u||[(||z_k - u||\mu/2 - ||f'(u)||]$, showing that $f(z_k) \geq f(x_0)$, for large k, whence S must be bounded. We now show that the sequence $\{x_k\}$ is Cauchy. Again by Taylor's theorem if $s > k$, $f(x_s) - f(x_k) \geq [f'(x_k), x_s - x_k] + \mu||x_s - x_k||^2/2$. Since S is bounded, $||x_s - x_k|| \leq D$ where D is the diameter of S. Thus $||x_s - x_k||^2 \leq \frac{2}{\mu}\{f(x_s) - f(x_k) + D||f'(x_k)||\}$ which shows that $\{x_s\}$ is a Cauchy sequence. By the completeness of E $\{x_s\}$ has a limit, say z, in E, and $f'(z) = 0$. If z is not unique, then $f'(z_1) = f'(z_2) = 0$, $z_1 \neq z$. Thus $f(z_1) - f(z_2) \geq \frac{\mu}{2}||z_1 - z_2||^2 \leq f(z_2) - f(z_1)$, a contradiction. Hence z is unique and is a minimizer of f.

<u>Remarks</u>: Useful remarks may be found in [1], [3] and [9].

II. NEWTONIAN STEPS AND ACCELERATION

Suppose that at the given point x_0, the function f satisfies the conditions of the first example, namely $\phi(x) = f''_{-1}(x_0)f'(x)$, where $f''_{-1}(x_0) = [f''(x_0)]^{-1}$. The corresponding iteration is $x_{n+1} = x_n - \rho_n f''_{-1}(x_0)f'(x_n)$. This algorithm, when $\rho_n \equiv 1$ is known as the "modified" Newton's method (see [3], p. 259, or [7], p. 696). In a similar manner if $f''_{-1}(x)$ exists and is uniformly bounded below on S, we may define $\phi(x_n) = f''_{-1}(x_n)f'(x_n)$. We shall do this below. It is clear from what has already been said that ϕ satisfies hypotheses of the above theorem. Our object now is to formulate an algorithm using $f''_{-1}(x_n)f'(x_n) = \phi(x_n)$ which will converge at a superliner rate.

In the following we set $\Delta(x,\rho) = f(x) - f(x - \rho f''_{-1}(x)f'(x))$ and $g(x,\rho) = \Delta(x,\rho)/\rho[f''_{-1}(x)f'(x),\ f'(x)]$.

Theorem. Assume the level set S is a convex subset of a Banach space E. For each x in S assume the F-derivative f'' is continuous on S, $f''(x)$ is onto, $||f''(x)|| \leq M$, and $[f''(x)z,z] \geq m||z||^2$ for some $m > 0$ and all z in E. Set $x_{k+1} = x_k - \rho_k f''_{-1}(x_k)f'(x_k)$, where ρ_k is chosen so that for $\theta < 1/2$, $0 < \theta \leq g(x_k,\rho_k) \leq 1 - \theta$ with $\rho_k = 1$ if possible. Then:

(a) There exists a number N such that if $k > N$ then $\rho_k = 1$.

(b) There is a unique minimizer of f and the sequence $\{x_k\}$ converges to it faster than any geometric progression.

Proof. We have for all x in S that $M||z||^2 \geq [f''(x)x,x] \geq m||z||^2$ and $m^{-1}||y||^2 \geq [y,f''_{-1}(x)y] \geq mM^{-2}||y||^2$. Thus if $\phi(x) = f''_{-1}(x)f'(x)$, then $[f'(x),\phi(x)] \geq mM^{-2}||f'(x)||^2$, showing that ϕ satisfies the conditions of the above theorem. Since f'' is bounded on S, f' is Lipschitz continuous, by the mean value theorem. By (c) above $\{x_k\}$ converges to a unique minimizer of f.

Expand $\Delta(x,\rho)$ to two terms in the Taylor series with remainder $[f''(\xi)h,h)$, where $h = \rho f''_{-1}(x)f'(x)$. Set $f''(\xi) = f''(x) + f''(\xi) - f''(x)$. Ther $g(x,\rho) = 1 - \rho/2 - \rho[(f''(\xi) - f''(x))f''_{-1}(x)f'(x),f''_{-1}(x)f'(x)]/2[f'(x),f''_{-1}(x)f'(x)]$ $\geq 1 - \rho/2 - \rho||f''(\xi) - f''(x)||M^2/2m^3$. Thus $|g(x,\rho) - 1 + \rho/2| \leq \rho||f''(\xi) - f''(x)||M^2/2m$. Since $\xi(\rho_k)$ lies between x_k and x_{k+1} $x_0,\xi_0,x_1,\xi,\ldots$

is a Cauchy sequence; and it, together with its limit z, is a compactum. Consequently, on this compactum f'' is uniformly continuous, so that $\{||f''(\xi(\rho_k) - f''(x)||\}$ converges to 0, showing that the choice $\rho_k = 1$ is eventually feasible.

To prove (b) we write $x_{k+1} - z = x_k - z - \rho_k f''_{-1}(x_k)f'(x_k) =$ $x_k - z - \rho_k f''_{-1}(x_k)f''(x_k)(x_k - z) + \rho_k f''_{-1}(x_k)[f''(x_k)(x_k - z) - f'(x_k)]$.

Thus $||x_{k+1}-z|| = ||x_k-z-\rho_k(x_k-z)|| + \rho_k||f''_{-1}(x_k)||\,||f''(x_k)(x_k-z)-f'(x_k)||$.

Since f' is F-differentiable at x_k, $||f'(z) - f'(x_k) - f''(x_k)(z - x_k)||$ $< \varepsilon\ ||z - x_k||$. Thus $||x_{k+1} - z|| = (1 - \rho_k)||x_k - z|| + \rho_k m^{-1}\varepsilon||z - x_k||$ Q.E.D.

Remarks:

(1) Both sides of the inverse of $f''_{-1}(x)$ are used in the proof.

(2) The analogue of the modified Newton process, namely choosing $\phi(x) = f''_{-1}(x_0)f'(x)$ or $f''_{-1}(x_k)f'(x)$ with k fixed also will under the hypothesis of the above theorem generate a sequence converging to a unique minimizer of f. Since $||x_{k+1}-z|| = ||x_k-z-\rho_k f''_{-1}(x_0)f''(z)(x_k-z)|| + \rho_k||[f''_{-1}(x_0)]||\ \varepsilon$ $||x_k-z||m^{-1}$ when $||x_k-z|| < \delta$, the rate of convergence is eventually geometric provided $||I - \rho_k f''_{-1}(x_0)f''(z)|| < 1$. Since

$||I - \rho_k f''_{-1}(x_0)f''(z)|| \leq 1 - \rho_k + \rho_k||f''_{-1}(x_0)||\ ||(f''(x_0) - f''(z)||$, if $||f''(x_0) - f''(z)||$ is sufficiently small, $\rho_k \equiv 1$ will generate a sequence converging to z at the rate of geometric progression. A sufficient condition for the global geometric convergence would be $(M/m) < 1/2$, since $||f''(x)|| \leq M$ and $||f''_{-1}(x)|| \leq m^{-1}$.

(3) Pertinent remarks may be found in [3].

III. EXAMPLE

a) We consider the following problem which arises from a linearized rendezvous problem. See for example [8], [9] and [4]. In [4] this problem is solved in the "dual". We consider here a construction in the "primal". In [8] and [9], we have discussed this problem in the spaces $\mathcal{L}_1$ and $\mathcal{L}_2$; we now discuss the problem in $\mathcal{L}_p$ for $p > 2$. Let $\mathcal{L}_p$ denote the direct sum of n $L_p[0,1]$ spaces. Thus a point $x \in \mathcal{L}_p$ if $x = (x_1,\ldots,x_n)$ and $x_i \in L_p[0,1]$; the norm in $\mathcal{L}_p$ will be $||x||_p = [\int_0^1 |x(t)|^p dt]^{1/p}$ where $|x(t)| = \left[\sum_{i=1}^n x_i^2(t)\right]^{\frac{1}{2}}$. Since $\sqrt{n}\max\{x_i(t) : 1 \le i \le n\} \ge |x(t)|$, $||x||_p$ is well defined. Let $\{u^i : 1 \le i \le m\}$ be a linearly independent set in $\mathcal{L}_p$. Set $\frac{1}{p} + \frac{1}{q} = 1$. Since $q < p$, u^i is also in $\mathcal{L}_q$. Given numbers α_i, $(1 \le i \le m)$ define the affine subspace $M = \{x \in \mathcal{L}_p : [u^i,x] = \alpha_i : 1 \le i \le m\}$. We shall consider the problem of minimizing $f(x) = ||x||_p^p$ on M. The limits $p \to 1$ and $p \to \infty$ correspond to the cases of rendezvous with minimum fuel and minimum thrust amplitude respectively. In what follows we shall assume for simplicity that $n = 2$. There are no further difficulties in the general case.

We first observe that if the Gateaux differential (G-differential) of f exists it is given by:

$$f'(x)h = p\int_0^1 |x(t)|^{p-2}[x_1(t)h_1(t) + x_2(t)h_2(t)]dt$$

$$= p\int_0^1 |x(t)|^{p-1}\left[\frac{x_1(t)}{|x(t)|}h_1(t) + \frac{x_2(t)}{|x(t)|}h_2(t)\right]dt$$

$$\leq p||x||_p^{p/q}[||h_1||_p + ||h_2||_p].$$

Here Holder's inequality has been employed on the function $t \to |x(t)|^{p-1}$ which belongs to $L_q[0,1]$. We have also used $||\cdot||_p$ for the norm in $L_p[0,1]$. Thus the G-derivative of f exists.

Observe now that if the second G-differential exists it is given by:

$$[f''(x)h,k] = p(p-2)\int_0^1 |x(t)|^{p-4}(x_1(t)h_1(t) + x_2(t)h_2(t))(x_1(t)k_1(t) + x_2(t)k_2(t))dt$$

$$+ p\int_0^1 |x(t)|^{p-2}(k_1(t)h_1(t) + k_2(t)h_2(t))dt$$

$$= p(p-2)\int_0^1 |x(t)|^{p-2}\left(\frac{x_2(t)}{|x(t)|}h_1(t) + \frac{x_2(t)}{|x(t)|}h_2(t)\right)\left(\frac{x_1(t)}{|x(t)|}k_1(t) + \frac{x_2(t)}{|x(t)|}k_2(t)\right)dt$$

$$+ p\int_0^1 |x(t)|^{p-2}(k_1(t)h_1(t) + k_2(t)h_2(t))dt$$

$$\leq 2p(p-2)\int_0^1 |x(t)|^{p-2}|h(t)||k(t)|dt + p\int_0^1 |x(t)|^{p-2}|h(t)||k(t)|dt.$$

If $x \in L_p$, $y \in L_q$ and $z \in L_r$, and $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$, then $\int_0^1 |x(t)y(t)z(t)|dt \leq ||x||_p||y||_q||z||_r$. Since the function $t \to |x(t)|^{p-2}$ belongs to $L_{p'}$, where $p' = \frac{p}{p-2}$, and $\frac{1}{p'} + \frac{2}{p} = 1$ it follows that if $u(t) = |x(t)|^{p-2}$, $\left[\int_0^1 |u(t)|^{p'}dt\right]^{1/p'} = ||x||_p^{p-2}$, and $[f''(x)h,k] \leq (2p^2 - 3p)||x||_p^{p-2}||h||_p||k||_p$.

As before, let S denote the level set of f at x_0 where x_0 will be subsequently chosen in M. Thus if $x \in S$, $||x||_p^{p-2} = [f(x)]^{\frac{p-2}{2}} \leq [f(x_0)]^{\frac{p-2}{2}}$ showing that $[f''(x)h,k]$ is uniformly bounded on S, if h and k are confined to the unit sphere. It follows by Taylor's theorem that f' is F-differentiable on S. By the generalized mean value theorem it further follows that f' is Lipschitz continuous on S.

We now construct x_0 on M. Let $x_0 = \Sigma c_j u^j$. Thus x_0 lies on M if and only if $\sum\limits_j c_j[u^i,u^j] = \alpha_i$. We show that the null space of the matrix $\{[u^i,u^j]\}$ consists only of the 0 element so that c_i is uniquely determined. If for some $c_j \neq 0$, $\sum\limits_j c_j[u^i,u^j] = 0$ then $\sum\limits_i c_i \sum\limits_j c_j [u^i,u^j] = [\Sigma c_i u^i, \Sigma c_j u^j] = 0$, contradicting the linear independence of the set $\{u^i : 1 \leq i \leq n\}$. Let $N = \{x \in \mathcal{L}_p : [u^i,x] = 0,\ i=1,\ldots,m\}$. We now choose h to maximize $[f'(x),h]$ subject to $||h||_p = 1$ and $h \in N$. The maximum is achieved because the sphere meets N in a weakly compact set and the linear function $[f'(x),\cdot]$ is weakly continuous. The maximization can be accomplished by the method of Euler multipliers [10]. Let $\phi(h) = ||h||_p^p - 1$ and $\phi_i(h) = [u^i,h]$. Then a necessary condition that h maximize $f'(x,h)$ subject to $\phi(h) = \phi_i(h) = 0$ is that there exists c_i, $(1 \leq i \leq m+1)$, such that $f'(x)k =$

$$c_1 p \int_0^1 |h(t)|^{p-2}(h_1(t)k_1(t) + h_2(t)k_2(t))dt + \sum_{j=2}^{m+1} c_j[u^{j-1},k] \quad \text{for all} \quad k \in \mathcal{L}_p.$$

It follows that

$$p|x(t)|^{p-2}x_i(t) = pc_1|h(t)|^{p-2}h_i(t) + c_2u_i^1(t) + \ldots + c_{m+1}u_i^m(t) \quad i=1,2.$$

Let $f_i(t) = (pc_1)^{-1}[p|x(t)|^{p-2}x_i(t) - c_2u_i^1(t) - \ldots - c_{m+1}u_i^m(t)]$, and observe that

$$|h(t)|^{2p-2} = f_1^2(t) + f_2^2(t).$$

Therefore:

$$h_i(t) = [(f_1^2(t) + f_2^2(t)^{\frac{1}{2}}]^{\frac{q}{p}} f_i(t)/(f_1^2(t) + f_2^2(t))^{\frac{1}{2}}$$

showing that $h_i \in L_p$. We now solve the non-linear equations $\phi(h) = 0$, $\phi_i(h) = 0$ $(1 \leq i \leq m)$ for $c_2, \ldots, c_{m+1}$. If necessary, we replace h by $-h$ to ensure that we have a maximizer rather than a minimizer. The solution is now unique, due to the strict convexity of the sphere in $\mathcal{L}_p$. Because of the uniqueness of the extremal, the h we have constructed must be this extremal.

The subspace N is also an $\mathcal{L}_p$ space. Minimizing $f(x)$ on M is equivalent to minimizing $f(y + x_0)$ on N, with $x = y + x_0$. Clearly the gradient of the function f restricted to N is $h[f'(x),h]$. (See I2, above.) It follows, therefore, if $\phi(x) = h[f'(x),h]$, then ϕ satisfies the conditions required for the theorem of I.

By the theorem of I we may infer that every weak cluster point z of the sequence $\{x^k\}$ minimizes f and $f(x^k)$ converges downward to $f(z)$. Furthermore, since z is unique in the above problem, $\{x^k\} \rightharpoonup z$. We now show, moreover, that $\{x^k\} \to z$. For each component x^i of x, $(1 \leq i \leq n)$ we have that $0 \leq \int_0^1 |x^i(t)|^p dt \leq \int_0^1 |x(t)|^p dt$. Since $\{f(x_k)\}$ converges, the numbers $y_k^i = \int_0^1 |x_k^i(t)|^p dt$ are bounded and the sequence $\{x_k^i\}$ has a weak cluster point. In fact, $\{y_k^i\}$ converges. To prove this, observe that since $\{x_k\} \rightharpoonup z$, $\{x_k^i\} \rightharpoonup z_i$. Take a subsequence $\{x_k^i\}$ such that $\{y_k^i\} \to y^i$. Since $\{x_k^i\}$ converges both weakly and in L_p norm, $\{x_k^i\}$ converges strongly. Thus $\{x_k\} \to z'$, say. By continuity $f(x_k) \to f(z') = f(z)$. By uniqueness, $z = z'$. Therefore, the subsequence $\{x_k^i\} \to z^i$. Suppose that $\{y_k^i\}$ had another cluster point $\bar{y}^i \neq y^i$.

Take a new subsequence $\{x_k^i\}$ such that $\{v_k^i\} \to \bar{y}^i$. Again $\{x_k^i\} \to z^i$, and therefore, $\{v_k^i\} \to \int_0^T |z^i(\cdot)|^p dt = v^i$, a contradiction. It follows, therefore, that $\{x_k\} \to z$.

(b) The above processes require that at each cycle a non-linear system be solved to determine the gradient. This can be circumvented by imbedding the problem into a Hilbert space. Specifically, assume that the components of u^i are bounded and measurable. Let $\mathcal{L}_2$ denote the direct sum of $L_2[0,1]$ analogously to the above, and define $M' = \{x \in \mathcal{L}_2 : [u^i,x] = c^i : 1 \leq i \leq m\}$. Let f now be defined on M'. Since f achieves a minimum on M and $M \subset M'$, f also achieves a minimum on M'. Because M is dense in M', the minima are equal. The gradient of f on M' is merely the restriction of the gradient of f in $\mathcal{L}_2$ to M' and is obtained by orthogonal projection. See [9]. In general, $f'(x)$ does not exist. But if x is bounded and measurable, i.e., $x \in \mathcal{L}_\infty$, $f'(x) \in \mathcal{L}_2^*$ and $\nabla f(x) \in \mathcal{L}_2$. The set S is bounded in $\mathcal{L}_p$ and this implies S is bounded in $\mathcal{L}_2$, since $||x||_2 \leq ||x||_p$ if $p \geq 2$.

Since f is convex and continuous, S is closed bounded and convex; furthermore, the derivatives of f are densely defined on S. Assume $x_n \in M'$, $x_n \in \mathcal{L}_\infty$ and $u^i \in \mathcal{L}_\infty$, $(1 \leq i \leq m)$. Then x_{n+1} is well defined and is also in $\mathcal{L}_\infty$. To see this, verify that $\nabla f(x_n) \in \mathcal{L}_\infty$ and the projection of $\nabla f(x_n)$ on the set $\{x \in \mathcal{L}_2 : [u^i,x] = 0, (1 \leq i \leq n)\}$ is also in $\mathcal{L}_\infty$. We are able to conclude again therefore that there is a unique minimizer z, and that $\{x_k\} \to z$ and $f(x_k) \searrow f(z)$.

REFERENCES

[1] Goldstein, A. A. "On steepest descent", Boeing document D1-82-0413, Seattle, Washington, February, 1965; *SIAM Control Series A*, Vol. 3, No. 1.

[2] Goldstein, A. A. "On Newton's method", Boeing document D1-82-0426, Seattle, Washington, April, 1965.

[3] Vainberg, M. M. "On the convergence of the method of steepest descents for non-linear equations", *Amer. Math. Soc. Doklady Translations* 1, 1 (1960).

[4] Neustadt, L. W. "Optimization, a moment problem and nonlinear programming", *Journ. SIAM Control*, Vol. 2, No. 1, 1964.

[5] Vainberg, M. M. Variational Methods for the Study of Non-linear Operators, Holden-Day, Inc., 1964.

[6] Riesz, F. and Sz-Nagy, B. Lecons d'Analyse Fonctionelle.

[7] Kantorovich, L. V. and Akilov, G. P. Functional Analysis in Normed Spaces, Macmillan, 1964.

[8] Goldstein, A. A. "Fuel optimization in orbital rendezvous", *Proceedings of the AIAA Conference on Guidance and Control*, Cambridge, Mass., August, 1963, with A. Green, A. Johnson, in Vol. 13 *Progress in Astronautics*, Academic Press, April, 1964, pp. 823-844.

[9] Goldstein, A. A. "Minimizing functionals on Hilbert space", appeared Computing Methods in Optimization Problems, Academic Press, April, 1964, pp. 159-166.

[10] Lusternik, L. A. "On conditional extrema of functionals", *Mat. Sb.* 41, 3, (1934), pp. 390-401. See also L. A. Lusternik and V. J. Sobolev, Elements of Functional Analysis, Ungar N. Y., 1961, p. 210.

ACKNOWLEDGMENTS

This work was supported by the Boeing Scientific Research Laboratories and by U. S. Air Force grant AF-AFOSR-937-65.